REAL-TIME DISPARITY MAP EXTRACTION IN A DUAL HEAD STEREO VISION SYSTEM

G. CALIN† and V. O. RODA†

† Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, Brazil
gcalin@uol.com.br, valentin@sel.eesc.usp.br

Abstract — This paper describes the design of an algorithm for constructing dense disparity maps using the image streams from two CMOS camera sensors. The proposed algorithm extracts information from the images based on correlation and uses the epipolar constraint. For real-time performance, the processing structure of the algorithm was built targeting implementation on programmable logic, where pipelined structures and condensed logic blocks were used.

Keywords — Stereo vision, disparity map, programmable logic.

I. INTRODUCTION

Researchers have been giving especial attention to computer vision systems capable of delivering accurate 3D information of an observed scene, which leads to the construction of robust intelligent vehicles. Using low cost sensors, it has been possible to develop stereo vision systems capable of extracting 3D features by passive sensing of the environment.

Most stereo vision implementations are based on a two camera configuration setup, where each camera delivers a two 2D representation of a given scene, as show in Fig. 1. Stereo vision is achieved by extracting 3D information by processing two or more 2D images of a given scene. The processing for extracting the 3D information creates a map that describes which point in the 2D images corresponds to the same point in the 3D scene. Detailed description of the stereo vision problem has been widely studied in past and it is not presented. Refer, for instance, to Grosso et al. (1989) and Grosso and Tistarelli (1995) for a detailed study of this subject.

Several stereo algorithms have been proposed in recent years to solve the problem of finding the correspondence of the right and left image. Simple methods employ the measure of absolute or squared differences of the pixels intensities, to measure the similarity between the images (Suniyoto et al., 2004). Other methods, in order to increase accuracy, employ window-based matching, where a cost function is evaluated around the pixel of interest to find the best match. These methods usually do not consider occlusions and present problems in regions displaying little or repetitive textures, leading to similar cost functions and being unable to find the proper match (Darabiha et al., 2003; Silva et al., 2003; Cox et al., 1996).

Birchfield and Tomasi (1998a, b) employed dynamic programming to solve the matching problem, where each scanline —and in some cases in-between scanlines— are described as a dynamic cost function and evaluated with addition of some penalties criteria, like occlusions and large jumps in disparity.

![Figure 1. Typical Stereo Vision problem: two cameras, acquiring images of the same scene, have two different 2D representation of a common 3D point. With proper processing, the position and depth of the 3D point can be extracted from the images.](image-url)